# 基于多尺度池化卷积神经网络的疲劳检测方法研究 *

顾王欢，朱 煜†，陈旭东，郑兵兵，何林飞

(华东理工大学 电子与通信工程系，上海 200237)

**摘 要：**疲劳驾驶是造成交通事故的主要原因之一，通过视觉传感器采集驾驶员面部图像并进行基于视觉特征分析的疲劳驾驶检测具有实际意义。针对视觉特征分析疲劳检测问题，设计了一种级联深度学习的检测系统结构，并提出基于多尺度池化的卷积神经网络疲劳状态检测模型。首先通过深度学习模型 MTCNN 进行人脸检测，提取出眼睛和嘴巴区域。针对眼睛和嘴巴的状态表征和识别问题，提出一种基于 ResNet 的多尺度池化模型（MSP）对眼睛和嘴巴状态进行训练。实时检测时，将眼睛嘴巴区域通过训练好的卷积神经网络模型进行状态识别，最后基于 PERCLOS 和提出的嘴巴张合频率(FOM)对驾驶员进行疲劳判定。实验结果表明，该算法具有较高的检测准确率，同时满足实时性要求，且对复杂环境具有较高的鲁棒性。

**关键词：**视觉特征分析；多尺度池化；卷积神经网络；疲劳检测；人脸检测

**中图分类号：**TP391.4 doi: 10.3969/j.issn.1001-3695.2018.05.0337

## Driver's fatigue detection system based on multi-scale pooling convolutional neural networks

Gu Wanghuan, Zhu Yu†, Chen Xudong, Zheng Bingbing, He Linfei

(*Dept. of Electronics & Communication Engineering, East China University of Science & Technology, Shanghai* 200237, *China*)

**Abstract:** Fatigue driving is one of the main causes of traffic accidents, particularly for drivers of large vehicles like buses and heavy trucks. For designing driver fatigue monitoring systems, the visual features based techniques is one of the most effective approaches. This paper proposed a hierarchical convolutional neural network model with multi-scale pooling for vision-based fatigue detection system. The first step is face detection and extraction of eye and mouse regions by deep learning model—MTCNN. In order to solve the problem of characterization and recognition of eye and mouth regions, this paper proposed a multi-scale pooling model（MSP）based on ResNet to train different states of eye and mouth. In real-time detection, the paper recognized the states of eye and mouth by the pre-trained convolutional neural network model. Finally, this paper detect fatigue through the PERCLOS and the frequency of open mouth (FOM) . The experimental results show that the proposed algorithm has high detection accuracy, real-time performance and high robustness to complex environments.

**Key words:** visual feature analysis; multi-scale pooling; convolutional neural network; fatigue detection; face detection

## 0 引言

美国汽车协会交通安全基金会(AAA Foundation for Traffic Safety) [1]的一项调查表明：16%-21%的交通事故是由于驾驶员疲劳驾驶导致的。尤其对于驾驶时间较长的卡车、公共汽车等大型车辆尤为严重。因此，驾驶员疲劳检测技术对于预防交通事故有着重要的研究意义。

当人进入疲劳状态时，往往会伴随各种生物特征上的变化，如打哈欠、长时间的闭眼、脑电波和心电波的变化等。本文针对需求将疲劳检测的研究主要集中于对驾驶员面部行为特征的分析上。基于驾驶员面部行为特征的疲劳检测算法一般分为两步:(a)人脸检测；(b)根据面部行为状态判断疲劳。

针对人脸检测算法的研究有很多，文献[2]是 Vlola 的经典之作，该文首次引入积分图概念，结合 Adaboost 算法进行人脸检测，在保证检测效果的同时提高了检测速度，做到了实时监测；Yong Du 等人[3]提出一种基于驾驶员眼睛变化状态判断疲劳的方法，该方法使用了帧差法绑定颜色信息检测人脸；2012年卷积神经网络(convolutional neural networks,CNN) [4]在ImageNet 大赛上一鸣惊人，使得深度学习(deep learning, DL)成为近年来研究图像视频等邻域最热门的方法，如图像检索 [5]、

**收稿日期：**2018-05-03；**修回日期：**2018-06-04 **基金项目：**上海市科学技术委员会科研计划项目（17DZ1100808，17DZ1100803）

**作者简介：**顾王欢（1994-），男，硕士研究生，主要研究方向为图像处理、深度学习；朱煜（1973-），女（通信作者），教授，博士，主要研究方向为图像与视频处理、机器学习（zhuyu@ecust.edu.cn）；陈旭东（1993-），男，硕士研究生，主要研究方向为计算机视觉；郑兵兵（1992-），男，博士研究生，主要研究方向为图像处理、深度学习；何林飞（1994-），男，硕士研究生，主要研究方向为计算机视觉、机器学习．

人体行为识别[6]、场景分析[7]等均取得了巨大的成功。在人脸检测和识别[8]的应用中，深度学习方法亦是效果显著。Taigman 等人[9]在 2014 年提出了 DeepFace 模型，将深度学习模型成功应用于人脸检测和识别中，解决了传统方法中普遍存在的鲁棒性差的问题，大大提高了人脸检测和识别的准确性；2016 年，Zhang 等人[10]提出了一种多任务的卷积神经网络模型(multitask convolutional neural networks, MTCNN)，将人脸检测分为三层网络进行训练。该模型能够快速准确地检测多个人脸，并对侧脸检测具有良好的鲁棒性，其在 FDDB、WIDERFACE 和 AFLW 人脸数据集上的验证结果达到 95%。

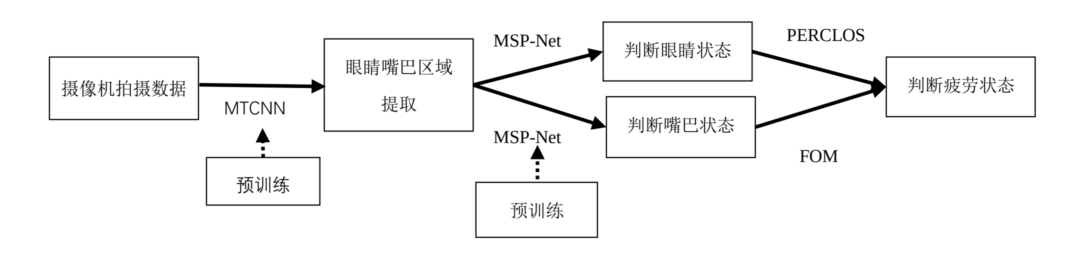



图 1　算法流程

针对面部状态判断疲劳的方法主要是从眼睛的睁闭和打哈欠的行为进行判断的。Wierwille 等人[11]建立了一种 PERCLOS 的理论模型，该模型将 PERCLOS 定义为单位时间内眼睑闭合程度达到 70%或者 80%的时间占比，PERCLOS 已成为一种非常有效的疲劳判定的指标；Chu 等人[12]则通过研究嘴巴的形状，将嘴巴的数据输入到 BP 网络中进行训练，分类出三种不同精神状态的类别。

用深度学习进行疲劳检测已成为近年来的热点之一。赵雪鹏等人[13]使用了一种级联网络来对眼睛部位进行定位和状态检测，以此判断疲劳；耿磊等人[14]则是首先使用 AdaBoost 和 KCF (kernelized correlation filters)进行人脸的检测和跟踪，然后再使用经典的网络结构对眼睛和嘴巴的状态进行检测。

基于视觉图像处理的疲劳检测是一个已经研究多年的课题，然而由于真实场景下环境光照、人员表情情况干扰较大，在人脸检测精度和面部行为特征检测精度上依然有待提高。针对此问题，本文提出了一种级联深度学习结构及基于多尺度池化卷积神经网络(multi-scale pooling networks, MSP-net)的疲劳状态检测模型。通过已经训练好的 MTCNN 模型检测出人脸，提取眼睛和嘴巴的数据，将眼睛和嘴巴数据分别放入已经训练好的 MSP-Net CNN 网络中进行检测，判断眼睛的睁闭和嘴巴的张合情况，最后根据 PERCLOS 和嘴巴张合频率(frequency of open mouth, FOM)联合判断疲劳。具体流程如图 1 所示。

## 1　预处理

### 1.1　数据采集

本文的疲劳检测系统采用两个深度学习模型级联形式，一个是针对检测人脸及提取眼睛和嘴巴图像的 MTCNN，另一个是本文提出的对眼睛和嘴巴图像进行状态判定的多尺度池化 CNN 网络。这两个网络都需要预训练。

由于近年来深度学习模型在人脸检测方面已经取得了骄人的成绩，本文利用已有的优秀网络来获得人脸及眼嘴区域，而将重点工作放在疲劳检测任务中。本文采用的 MTCNN 网络是由 KZhang 等人已经训练好的模型，作者使用 WIDER FACE 数据库[15]，人脸图像在 20 万张左右，包含边框标注和 5 个关键点信息（两个眼球的中心坐标、鼻尖坐标和两个嘴角坐标）。网络训练回归的是人脸边框和 5 个关键点坐标。

眼睛和嘴巴图像状态判定网络训练的数据由本文笔者在实际工作中自主采集，包括部分实际驾驶员图像及 21 位志愿者协助获得。考虑到驾驶员有夜间驾驶的情况，采集的设备除了使用普通摄像头以外，还使用红外摄像头。在图像和视频采集过程中，综合考虑驾驶员实际驾驶时的各种复杂环境问题，采集的图像数据包括眼睛睁闭、嘴巴张合、无眼镜、佩戴眼镜、正面和侧面等各种情况。采集之后进行筛选处理，剔除部分噪声样本。最后将筛选好的样本进行分类：普通摄像头的睁眼、闭眼、张嘴、闭嘴和红外摄像头的睁眼、闭眼、张嘴、闭嘴情况。其中，眼睛一共 36 764 个样本，嘴巴一共 15 185 个样本。样本示例如图 2 所示。

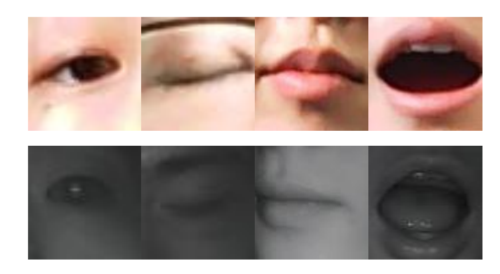


图 2　采集样本示例

### 1.2　数据预处理

由于红外摄像头采集的样本图像亮度普遍偏低，直接放入网络训练效果不好，所以，还需对样本进行预处理。本文使用直方图均衡[16]的方法。

直方图均衡的目的是为了增强局部对比度。其主要步骤是求出原图灰度级的累计概率，根据映射关系将原灰度值映射到新的灰度值。如灰度值为 0~255 的灰度图像，其灰度映射关系如下:

$$p_k = 255 * \sum_{i=0}^{k} \frac{n_i}{n} \quad k = 0,1,2 \ldots, L-1 \tag{1}$$

其中：$p_k$表示映射后的灰度值；$k$表示第$k$个灰度级；一共$L$个灰度级；$n$表示所有像素个数；$n_i$表示第$i$个灰度级的像素个数；得出的$p_k$最后再取整。图 3 展示了直方图均衡前后的对比。



图 3　直方图均衡前后对比

## 2　人脸检测

传统人脸检测算法，如 AdaBoost 或者帧差法等对于复杂环境的鲁棒性较差，深度学习模型在这方面有着其巨大的优势。本文借鉴 Zhang 等人的 MTCNN 模型进行人脸检测部分的工作。

MTCNN 的网络结构如图 4 所示，主要分为 P-NET、R-NET 和 O-NET 三层。

P-NET：首先构建图像金字塔；然后通过一个全卷积网络(fully convolutional network, FCN)[17]，获取候选脸部窗口(candidate facial windows)和边框回归向量(bounding box regression vectors)，用该边框回归对候选脸部窗口进行校准；接着使用非极大值抑制(nonmaximum suppression，NMS)[18]合并高度重合候选区域。

R-NET：将 P-NET 得到的候选区域放入此网络中，进一步筛查掉大量错误候选区和做校准，最后同样使用 NMS 进行合并候选区域。

O-NET：这一层网络与 R-NET 类似，但这一层做得更加细致，将对候选区域进行更严格的监督，最后多输出 5 个关键点坐标。

MTCNN 的鲁棒性很好，对于旋转了一定角度的人脸依然可以准确地检测。通过 MTCNN 检测人脸，并根据返回的 5 个关键点，可以成功标志出眼睛嘴巴部位。如图 5 所示，左图是普通摄像头下的检测结果，右图是红外摄像头下的检测结果。

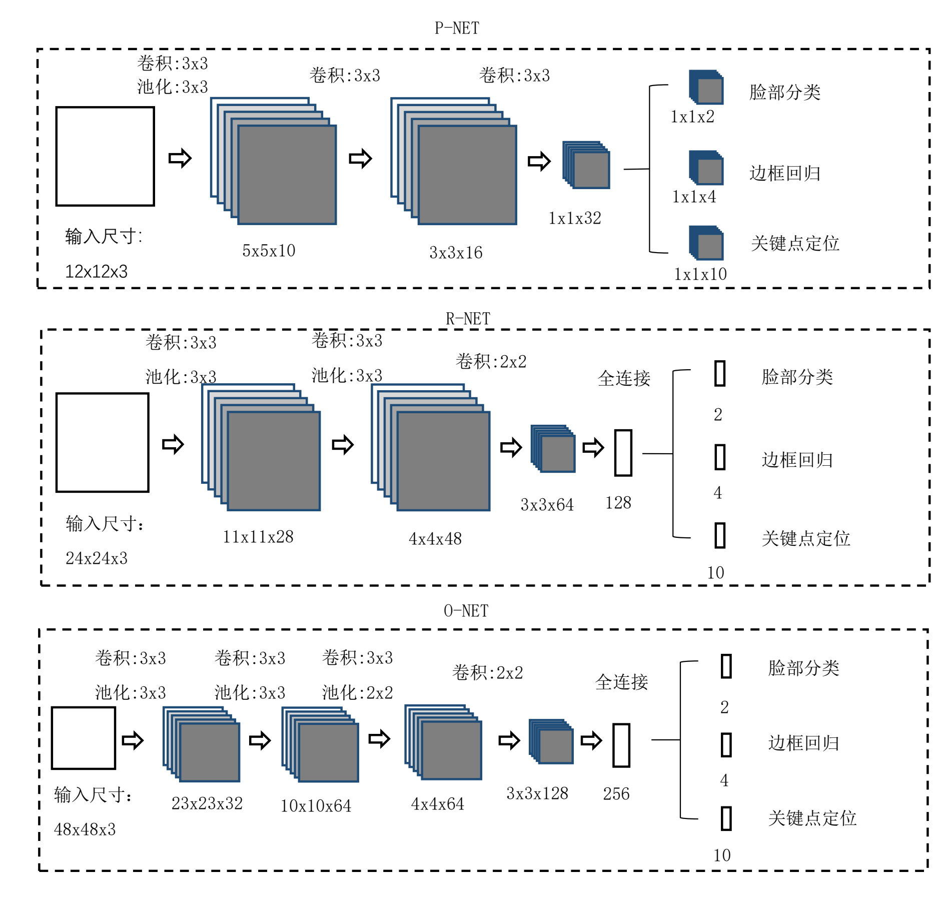



图 4　MTCNN 结构

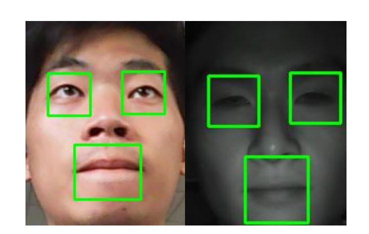


图 5　MTCNN 人脸检测与眼睛嘴巴区域提取

## 3　多尺度池化卷积神经网络 MSP-Net

提取出眼睛和嘴巴部位后，就需要进行眼睛和嘴巴的状态检测。本文提出了一种多尺度池化卷积神经网络模型(multi-scale pooling networks, MSP-net )分别训练眼睛和嘴巴的状态。

### 3.1　模型设计

多尺度池化卷积神经网络模型(MSP-net)在 ResNet[19]的结构基础上进行改进，保留了残差概念，在 pooling 层将原来的 max pooling 修改为本文提出的多尺度池化(MSP)，用以提高对不同分辨率下采集图像的识别效果。MSP 的结构如图 6 所示。

MSP 模块步骤如下：a）将上一级网络传入的特征图(feature maps)先经过两次 max pooling 得到一组边长比原特征图小 4 倍的特征图；b）对原特征图进行尺度缩放操作，边长缩小一倍，

再对得到的新特征图进行一次 max pooling 得到另一组边长为原特征图 1/4 大小的特征图；c）对原特征图再进行尺度缩放，边长缩小为原来的 1/4 倍；d）将三次输出的特征图进行级联，传入后面的深度学习网络。

多尺度池化模型的思想来源于空间金字塔池化模型[20]。相比于空间金字塔模型，多尺度池化的优点在于其代入位置更加灵活，可以在网络的开头、中间或是结尾多次使用。

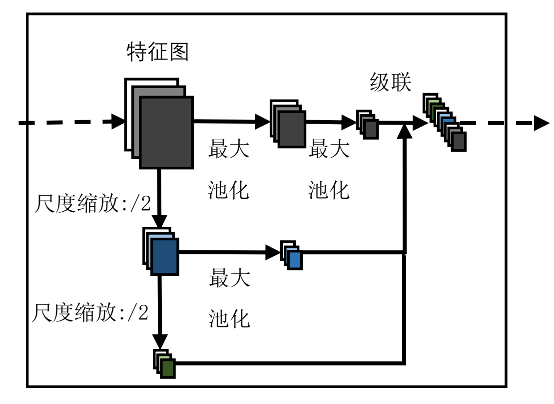



图 6　多尺度池化(msp)模块

本文设计的 MSP-net 网络的结构如图 7 所示。眼睛与嘴巴的训练方式一致，这里以眼睛为例。在 MSP-net 中，输入图像为 48*48 大小的灰度图，经过一次卷积和 MSP 输出 12*12*48 的 feature maps，卷积核大小为 3*3；之后经过一个残差块(residual block)，残差块中有两层卷积，卷积核大小还是 3*3，残差块输出 12*12*48 的 feature maps；再经过一次 max pooling，输出 6*6*48 的 feature maps；之后将 feature maps 转换成一维向量，进入全连接层，全连接层输入序列长为 1 728，有一层隐含层，长度为 1 000，最后经过 softmax 输出分类结果，类别分为普通摄像头睁眼、红外摄像头睁眼、普通摄像头闭眼和红外摄像头闭眼四类。

为了验证 MSP-net 的有效性，本文除了使用 MSP-net 进行实验，同时使用经典 AlexNet[4]和 ResNet 结构作为对比网络进行了训练和测试，进行了相关对比实验。6.2 节显示了三种网络的对比实验结果。

### 3.2　损失函数与优化方法

网络最后使用的是 softmax 分类方法，且分为四类，softmax 函数定义为

$$p_j=\frac{\exp(y_j')}{\sum_{k=0}^{3}\exp(y_k')},j=0,1,2,3 \tag{2}$$

其中：$p_j$表示为第$j$类的概率；$y_j'=\sum_i h_i\cdot w_{i,j}+b_j$表示全连接层最后一层输出，$h_i$是上一层输出，$w_{i,j}$和$b_j$分别为最后一层的权重和偏置。

损失函数定义为交叉熵：

$$\mathrm{L}_m=-\sum_{j=0}^{3}1\{y=j\}logp_j \tag{3}$$

其中：$\mathrm{L}_m$表示第$m$个样本的交叉熵；$1\{y=j\}$表示示性函数，即当$y=j$时，该函数为 1，当$y\neq j$时，该函数为 0。式（3）是单个样本的损失函数，当有M个训练样本时，损失函数需要求平均：

$$L=\frac{1}{M}\sum_{m=1}^{M}L_m \tag{4}$$

优化方法使用了自适应矩估计方法(adaptive moment estimation, Adam)[21]。

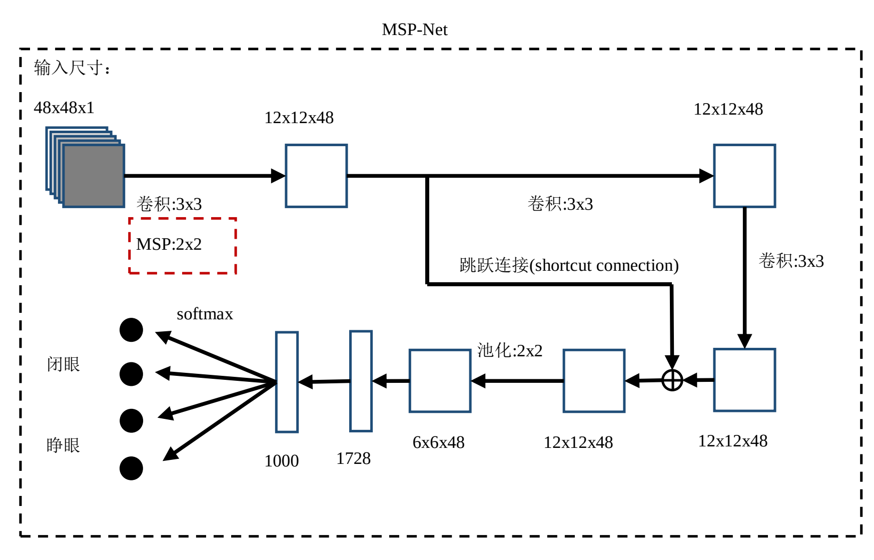



图 7　MSP-net 结构

## 4　疲劳状态检测

驾驶员在进入疲劳时会有一系列生物特征上的反应，如长时间闭眼、打哈欠等。基于这种生物特征的反应和之前网络获取的眼睛和嘴巴状态，通过计算 PERCLOS 和嘴巴张合频率(FOM)对驾驶员疲劳程度进行标定。

### 4.1　PERCLOS

PERCLOS 表示的是单位时间内眼睛闭合帧数和单位时间总帧数之间的比值[11]。其计算公式如下：

$$f_{per}=\frac{n}{N}\times 100 \tag{5}$$

其中：$n$表示闭眼帧数；$N$表示单位时间总帧数。PERCLOS 可以很好地量化驾驶员闭眼的程度，当 PERCLOS 达到某一个阈值时（文献[11]给出 0.15），可以判断驾驶员闭眼时长过长，已

经可以初步认为其进入了疲劳状态。

### 4.2　嘴巴张合频率（FOM）

嘴巴张合频率(FOM)与 PERCLOS 类似，表示的是单位时间内嘴巴张着的帧数和单位时间总帧数之间的比值。其计算公式如下：

$$f_{fom} = \frac{n}{N} \times 100 \tag{6}$$

与 PERCLOS 一样，$n$表示闭嘴帧数，$N$表示单位时间总帧数。且这两个指标都是值越大，表示疲劳程度越大。最后的疲劳状态检测需要联合两者同时考虑。

### 4.3　疲劳状态检测

所有深度网络的预训练完成以后，根据 PERCLOS 和 FOM 定好疲劳状态阈值，把整个网络系统应用到实时检测中。具体步骤为：摄像头拍摄驾驶员视频，MTCNN 抓取每一帧的人脸和 5 个关键点部位，提取出眼睛和嘴巴区域，利用 MSP-net 对每一帧抓取出来的眼睛和嘴巴部分进行状态检测，保存到固定长队列中，算法实时检测队列中值的变化情况，当队列中所有值的分布情况达到阈值疲劳状态时，报警机制就启动，提醒驾驶员已进入疲劳。

## 5　实验与分析

### 5.1　实验平台与数据集

所有的样本采集、网络训练和结果测试均是基于 Python3.6 和 Tensorflow1.2 实现的。摄像头使用普通摄像头和 920 nm 的红外摄像头。处理器为 CPU（3.20 GHz），内存 4 GB。眼睛和嘴巴的数据集(eye & mouth database, EMD)，是通过普通摄像头和红外摄像头一同采集，一共采集 21 位志愿者的数据，整理出眼睛 36 764 个样本，嘴巴 15 185 个样本。

为进一步检验提出的 MSP-Net 的检测效果，本文还针对人眼状态公共数据集 ZJU Eyeblink Database[22]数据集（图 8）进行了实验和算法性能对比。该数据库采集了 20 个人共 80 个视频剪辑。每个人 4 个剪辑：不带眼镜正视图剪辑、带薄边眼镜正视图剪辑、带黑边眼镜正视图剪辑和不带眼镜仰视图剪辑。该数据集图像分为睁眼和闭眼两类。睁眼样本共 7 000 个，其中训练样本 5 770 个，测试样本 1 230 个；闭眼样本共 1 984 个，其中训练样本 1 574 个，测试样本 410 个。每张图片大小 24*24。

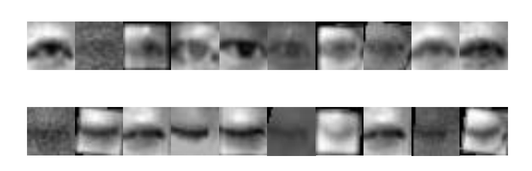


图 8　ZJU 数据集样本示例

### 5.2　实验结果分析

MTCNN 人脸检测效果显著。图 9 为 MTCNN 在本文的实际检测结果。从实验结果可以看到，即使测试者脸部转过一定角度，MTCNN 依然可以很好地检测到人脸。本文主要针对眼睛和嘴巴状态检测和疲劳状态检测进行相关实验。

#### 5.2.1 眼睛嘴巴状态检测

自建数据集 EMD 中眼睛样本 36 764 张，嘴巴样本 15 185 张，训练样本占 95%，测试样本占 5%。训练的具体分类和样本数如表 1 所示。训练前进行灰度化、归一化和直方图均衡等操作，输入网络的图像为 48*48 的灰度图。为了验证 MSP-Net 网络对眼睛嘴巴状态检测的效率，分别使用经典 AlexNet 和 ResNet 进行参数训练和测试，并作比较。各网络的测试结果如表 2 所示。

从实验结果可以看到，本文提出的 MSP-net 对于眼睛嘴巴状态的检测有着较高的准确率。此外，比较眼睛和嘴巴的实验结果发现，嘴巴的准确率普遍比眼睛的高，这是因为眼睛状态判别受到有带眼镜、不带眼镜及眼镜形态等干扰，所以分类效果比嘴巴的略低。实际检测效果如图 10 所示。图左上方为检测结果的文字描述。从实验效果可以看到，MSP-net 的鲁棒性很高，在佩戴眼镜、光线较弱且有转头的情况下，依然能够准确地识别出眼睛和嘴巴的状态。

疲劳检测除了需要较高的检测准确率，还需要快速检测的能力。因此，网络的检测速度也很重要。表 3 显示的是三种网络在 CPU 上检测一张 48*48 灰度图片的运算速度。

从表 3 可以看出，AlexNet 的运算速度要远远慢于 ResNet 和 MSP-net，原因在于 AlexNet 的参数量多很多，所以对于疲劳检测来说，AlexNet 并不适用。MSP-net 同 ResNet 在准确率和实时性上均比较出色，MSP-net 要稍优于 ResNet。

为了证明 MSP-Net 网络的普适性，本文在 ZJU 眼睛状态数据集上与其他两种方法做了比较实验。实验结果如表 4 所示。实验内容是判断眼睛的睁闭状态。表中，“Acc.”表示准确率，“AUC”表示的是受试者工作特征曲线(receiver operating characteristic curve, ROC) [23] 下的面积。“Acc.”及“AUC”值越大，表明模型性能越优。文献[24]使用了方向梯度直方图(histogram of oriented gradient, HOG) 描述子和随机蕨算法(random ferns)提取眼睛特征并检测睁闭；文献[22]则提出了一种新的特征描述子：主向梯度多尺度直方图（MultiHPOG），以提高对图像噪声和尺度变化的鲁棒性。从实验结果中可以看出，本文方法不管在准确度上还是耗时上均优于上述两种方法。

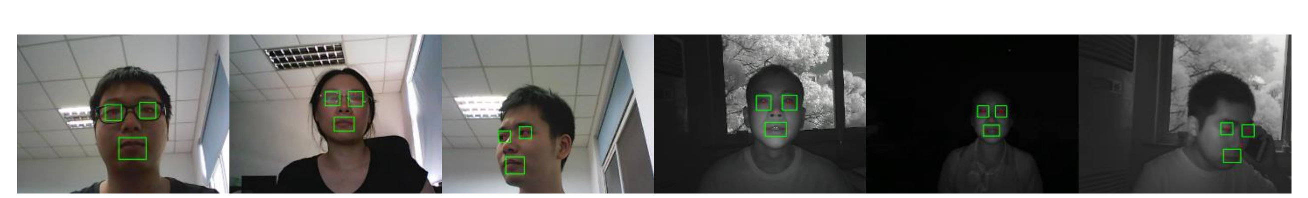


图 9　MTCNN 实际检测结果

表 1　眼睛嘴巴训练样本详细数据

| 数据集 | 训练对象 | 类别和训练样本数 | | | |
|---|---|---|---|---|---|
| EMD | 眼睛 | 普通摄像头睁眼 | 红外摄像头睁眼 | 普通摄像头闭眼 | 红外摄像头闭眼 |
| | | 10 034 | 5 381 | 12 600 | 7 126 |
| | 嘴巴 | 普通摄像头张嘴 | 红外摄像头张嘴 | 普通摄像头闭嘴 | 红外摄像头闭嘴 |
| | | 3 802 | 2 606 | 4 877 | 3 661 |

表 2　眼睛嘴巴状态检测测试结果对比

| 数据集 | 分类对象 | 测试样本数量 | 网络 | 准确率/% |
|---|---|---|---|---|
| EMD | 眼睛 | 1 853 | AlexNet | 97.192 |
| | | | ResNet | 97.786 |
| | | | MSP-Net | **98.056** |
| | 嘴巴 | 789 | AlexNet | 98.606 |
| | | | ResNet | 98.479 |
| | | | MSP-Net | **98.859** |




图 10　眼睛嘴巴实际判断情况

表 3　眼睛嘴巴状态检测模块单张图片各网络检测速度

| 网络 | 耗时/ms |
|---|---|
| AlexNet | 32.8 |
| ResNet | 5.6 |
| MSP-Net | **4.3** |

表 4　几种方法在 ZJU 数据集上的表现结果

| 方法 | ZJU 数据集 | | |
|---|---|---|---|
| | Acc./ % | AUC/ % | 耗时/ms |
| HOG + Random Ferns[24] | 94.76 | 98.89 | 9.000 |
| MultiHPOG + SVM[22] | 96.83 | 99.27 | 37.570 |
| MSP-Net（本文） | **96.89** | 99.05 | **1.823** |

表 5　疲劳检测实验结果

| | 疲劳次数 | 误检次数 | 漏检次数 | 查准率/ % | 查全率/ % |
|---|---|---|---|---|---|
| 普通摄像头 | 100 | 2 | 1 | 98.02 | 99.00 |
| 红外摄像头 | 100 | 5 | 2 | 95.15 | 98.00 |

#### 5.2.2 疲劳状态检测

Wierwille 等人[11]在其论文中给出：当进入疲劳时，PERCLOS 值大于 0.15，而其他研究者也有根据数据将 PERCLOS 阈值标定为其他值，如 0.25[12]。

本文在 PERCLOS 基础上同时考虑了 FOM 的影响，提出一种疲劳检测的方案：(a)当 PERCLOS 大于等于 0.5 或者 FOM 大于等于 0.5 时，判断为疲劳；(b)不满足(a)情况时，当 PERCLOS 大于等于 0.4 同时 FOM 大于等于 0.3 时，同样判断为疲劳。

疲劳检测系统实验中，拍摄 10 名志愿者模拟驾驶的过程，每人模拟 10 次疲劳状态，共 100 次疲劳。同时要求期间可以出现各种干扰状态，如增加眨眼频率、说话、轻笑等。检测系统同时用普通摄像头和红外摄像头实时检测志愿者的状态。表 5 记录了实验结果。

查准率的计算公式为

$$P = (n_f - n_{mis})/(n_f - n_{mis} + n_{err}) \quad (7)$$

其中：$P$为查准率；$n_f$为疲劳次数；$n_{err}$为误检次数；$n_{mis}$为漏检次数。查准率反映了误检概率，查准率越高，误检越低。

查全率的计算公式为

$$R = (n_f - n_{mis})/n_f \quad (8)$$

其中：$R$为查全率。查全率反映了漏检概率，查全率越高，漏检越低。

从实验结果可以看出，系统的查准率和查全率均较高。相比于模拟夜晚状态的红外图像来说，白天普通摄像头拍摄的 RGB 图像具有更高的查准率和查全率。

### 5.3　讨论

在实验过程中，遇到以下几个问题。

在采集样本的过程中，嘴巴图像的采集由于程序的设计，得到的样本大小是不统一的。为了更有效地进行训练，本文使用各向异性缩放原理[25]，将输入图片的大小均缩放到 48*48。

实际情况中，驾驶员会说话、轻笑等，这些情况下嘴巴也是张开的。不过，相比于打哈欠，这些情况张嘴的幅度要小很多。为了减小检测误差，在嘴巴样本进行分类时，将嘴巴微张的样本均归为闭合的类，只有张开幅度较大的样本才算张开的类。

前期，针对眼睛嘴巴状态训练，进行了二分类，即将普通摄像头和红外摄像头采集的样本融合到一起，只检测眼睛睁闭和嘴巴张合，然而效果不理想。可能的原因是普通摄像头拍摄的图像和红外摄像头拍摄的图像差异性较大，卷积神经网络在自动拟合特征时无法很好地同时满足两者的差异性。经过分析，最终本文将类别分成四类，即区分普通摄像头和红外摄像头的样本。

## 6　结束语

本文设计了一种级联深度学习结构及基于多尺度池化卷积神经网络的疲劳检测实时系统。首先通过 MTCNN 网络检测驾驶员的人脸，提取出眼睛和嘴巴关键位置，然后将眼睛和嘴巴图像送入多尺度池化 MSP-Net 中进行状态测试，设定固定长度队列，队列保存单位时间内对每一帧检测的结果，通过 PERCLOS 和嘴巴张合频率(FOM)联合判断疲劳状态。实验表明，本文提出的算法具有较高的检测准确率，同时可以达到实时检测的效果，且对复杂环境（如带眼镜、一定程度的偏头和黑夜驾驶等）有着较高的鲁棒性。本文提出的方法将进一步向嵌入式平台移植和优化。

## 参考文献：

[1] Tefft B C. Acute sleep deprivation and risk of motor vehicle crash involvement [R]. [S. l. ] : AAA Foundation for Traffic Safety, 2016.

[2] Viola P, Michael J J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57 (2): 137-154.

[3] Du Yong, Ma Peijun, Su Xiaohong, *et al.* Driver fatigue detection based on eye state analysis [J]. Technology & Health Care Official Journal of the European Society for Engineering & Medicine, 2008, 23 (s2): S453-S463.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. [S. l. ] : Curran Associates Inc, 2012: 1097-1105.

[5] 刘海龙，李宝安，吕学强，等. 基于深度卷积神经网络的图像检索算法研究 [J]. 计算机应用研究，2017, 34 (12): 3816-3819. (Liu Hailong, Li Baoan, Lyu Xueqiang, *et al.* Image retrieval based on deep convolutional neural network [J]. Application Research of Computers, 2017, 34 (12): 3816-3819. )

[6] 朱煜，赵江坤，王逸宁，等. 基于深度学习的人体行为识别算法综述 [J]. 自动化学报，2016, 42 (6): 848-857. (Zhu Yu, Zhao Jiangkun, Wang Yining, *et al.* A review of human action recognition based on deep learning [J]. Acta Automatica Sinica, 2016, 42 (6): 848-857. )

[7] Yu Tianshu, Wang Ruisheng. Scene parsing using graph matching on street-view data [M]. [S. l. ] : Elsevier Science Inc, 2016.

[8] Shao Hong, Chen Shuang, Zhao Jieyi, *et al.* Face recognition based on subset selection via metric learning on manifold [J]. Frontiers of Information Technology & Electronic Engineering, 2015, 16 (12): 1046-1058.

[9] Taigman Y, Yang Ming, Ranzato M, *et al.* DeepFace: closing the gap to human-level performance in face verification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l. ] : IEEE Computer Society, 2014: 1701-1708.

[10] Zhang Kaipeng, Zhang Zhanpeng, Li Zhifeng, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499-1503.

[11] Wierwille W W, Ellsworth L A. Evaluation of driver drowsiness by trained raters [J]. Accident Analysis & Prevention, 1994, 26 (5): 571-81.

[12] Chu Jiangwei, Jin Lisheng, Tong Bingliang, *et al.* A monitoring method of driver mouth behavior based on machine vision [C]// Proc of IEEE Intelligent Vehicles Symposium. 2004: 351-356.

[13] 赵雪鹏，孟春宁，冯明奎，等. 基于级联卷积神经网络的疲劳检测 [J]. 光电子·激光，2017 (5): 497-502. (Zhao Xuepeng, Meng Chunning, Feng Mingkui, *et al.* Fatigue detection based on cascade convolutional neural network [J]. Journal of Optoelectronics·laser, 2017, 28 (5): 497-502. )

[14] 耿磊，袁菲，肖志涛，等. 基于面部行为分析的驾驶员疲劳检测方法 [J]. 计算机工程，2018, 44 (1): 274-279. (Geng Lei, Yuan Fei, Xiao Zhitao, *et al.* Driver fatigue detection method based on facial behavior analysis [J]. Computer Engineering, 2018, 44 (1): 274-279. )

[15] Yang Shuo, Luo Ping, Chen C L, *et al.* WIDER FACE: a face detection benchmark [C]// Proc of Computer Vision and Pattern Recognition. [S. l. ] : IEEE Press, 2016: 5525-5533.

[16] Jaspers C A M. Histogram equalization: USA, US 6741736 B1 [P]. 2004-05-25.

[17] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39 (4): 640-651.

[18] Neubeck A, Gool L V. Efficient non-maximum suppression [C]// Proc of IEEE International Conference on Pattern Recognition. 2006: 850-855.

[19] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[20] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2014, 37 (9): 1904-1916.

[21] Kingma D P, Ba J. Adam: A method for stochastic optimization [C]// Proc

of International Conference on Learning Representations. 2015: 1-13.

[22] Song Fengyi, Tan Xiaoyang, Liu Xue, *et al*. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients [J]. Pattern Recognition, 2014, 47 (9): 2825-2838.

[23] Fawcett T. An introduction to ROC analysis [J]. Pattern Recognition Letter, 2006, 27 (8): 861-874.

[24] Dong Yanchao, Zhang Yan, Yue Jiguang, *et al*. Comparison of random forest, random ferns and support vector machine for eye state classification [J]. Multimedia Tools & Applications, 2016, 75 (19): 11763-11783.

[25] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of Computer Vision and Pattern Recognition. [S. l. ] : IEEE Press, 2014: 580-587.